

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-290608

(43)Date of publication of application : 19.10.2001

(51)Int.Cl.

G06F 3/06

G06F 12/16

(21)Application number : 2000-107489

(71)Applicant : HITACHI LTD

(22)Date of filing : 10.04.2000

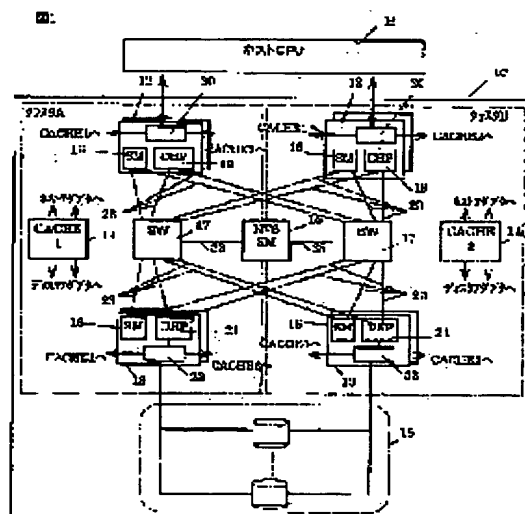
(72)Inventor : INOUE MITSURU

## (54) DISK CONTROLLER

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide a disk controller, which improves I/O performance, is immune to a power source fault and can be run easily.

**SOLUTION:** Concerning a multiprocessor control disk array controller 10 having plural processors 19 and 21, a shared memory for storing the control information of the respective processors 19 and 20 is composed of a nonvolatile shared memory 15 and a volatile shared memory 16. Thus, high I/O performance can be provided by making the shared memory versatile. By providing a nonvolatile shared memory plane, the power source fault can be coped with as well. By making only a part of the shared memory into nonvolatile shared memory plane, difficulties in circuit configuration or cost are reduced and running is facilitated.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office



## 【特許請求の範囲】

【請求項1】 複数のプロセッサを有するマルチプロセッサ制御ディスク制御装置であって、各プロセッサの制御情報を記憶する共有メモリを有し、該共有メモリは不揮発共有メモリ面と揮発共有メモリ面とからなることを特徴とするディスク制御装置。

【請求項2】 請求項1に記載のディスク制御装置において、前記各プロセッサは、前記共有メモリに書き込む情報を2重書き情報と1重書き情報とに分類し、前記2重書き情報は前記不揮発共有メモリ面と前記揮発共有メモリ面の両方に書き込み、前記1重書き情報は前記揮発共有メモリ面にのみ書き込むことを特徴とするディスク制御装置。

【請求項3】 請求項1または請求項2に記載のディスク制御装置において、前記不揮発共有メモリ面が独立した2以上の電源系のいずれからでも給電可能に配置され、前記揮発共有メモリ面が前記電源系の各々に分けて配置されることを特徴とするディスク制御装置。

【請求項4】 請求項1から請求項3のいずれかに記載のディスク制御装置において、前記不揮発共有メモリ面が、ディスクドライブを含むことを特徴とするディスク制御装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、ディスク制御装置に関し、さらに詳しくは、I/O性能が高く且つ電源障害に強く且つ容易に実施できるディスク制御装置に関する。

## 【0002】

【従来の技術】図8は、従来のディスク制御装置の一例を示す構成図である。このディスク制御装置1は、それぞれがプロセッサMPを有しホストコンピュータHとのインターフェースを行う複数のホストインターフェース部2、それぞれがプロセッサMPを有し複数のディスクドライブ3とのインターフェースを行う複数のディスクインターフェース部4、前記ディスクドライブ3に書き込むデータを一時記憶するキャッシュメモリ5(A)、5(B)、各プロセッサMPの制御情報を記憶する共有メモリ6(A)、6(B)、データ用共通バス7および制御情報用共通バス8を具備している。前記共有メモリ6(A)は、前記キャッシュメモリ5(A)に対応し、該キャッシュメモリ5(A)上のどの位置にどのデータが記憶されているかを示す情報を記憶する。また、前記共有メモリ6(B)は、前記キャッシュメモリ5(B)に対応し、該キャッシュメモリ5(B)上のどの位置にどのデータが記憶されているかを示す情報を記憶する。また、前記共有メモリ6(A)、6(B)には、ディスクドライブ3のハード構成等の情報が2重書きされる。さらに、これら共有メモリA面6(A)、6(B)は、電源障害に備え、バッテリーでバックアップされている。

【0003】他方、特開平10-333836号公報に開示のディスク制御装置では、共有メモリを多面化する(すなわち、2面以上とし且つ独立した複数のバスに分散配置する)ことで、I/O性能を向上している。

## 【0004】

【発明が解決しようとする課題】図8のディスク制御装置1では、バッテリーでバックアップされているため、電源障害時でも、共有メモリ6に記憶した情報が保護される。しかし、I/O処理要求数が増加すると、制御情報用共通バス8で通信が輻輳し、これがシステム性能のボトルネックになる問題点がある。

【0005】他方、特開平10-333836号公報に開示のディスク制御装置では、共有メモリが多面化されているため、I/O性能が高い。しかし、共有メモリがバッテリーでバックアップされていないため、電源障害時には、共有メモリに記憶した情報が失われてしまう問題点がある。

【0006】以上の従来技術を組み合わせ、特開平10-333836号公報の多面化された共有メモリをバッテリーでバックアップすれば、I/O性能を向上でき且つ電源障害にも強くなる、と考えられる。しかしながら、多面化された共有メモリを全てバッテリーでバックアップすることは、回路構成の複雑さやコストの観点から実施困難であり、現実的ではない。そこで、本発明の目的は、I/O性能が高く且つ電源障害に強く且つ容易に実施できるディスク制御装置を提供することにある。

## 【0007】

【課題を解決するための手段】第1の観点では、本発明は、複数のプロセッサを有するマルチプロセッサ制御ディスク制御装置であって、各プロセッサの制御情報を記憶する共有メモリを有し、該共有メモリは不揮発共有メモリ面と揮発共有メモリ面とからなることを特徴とするディスク制御装置を提供する。上記第1の観点によるディスク制御装置では、共有メモリを不揮発共有メモリ面と揮発共有メモリ面とに分けた構成とした。そして、揮発共有メモリの多面化によりI/O性能を向上でき、且つ、不揮発共有メモリ面を持つことで電源障害にも強くなる。さらに、揮発性共有メモリ面はバッテリーでバックアップしないから、回路構成やコスト上の困難が少なくなり、実施が容易になる。

【0008】第2の観点では、本発明は、上記構成のディスク制御装置において、前記各プロセッサは、前記共有メモリに書き込む情報を2重書き情報と1重書き情報とに分類し、前記2重書き情報は前記不揮発共有メモリ面と前記揮発共有メモリ面とに2重に書き込み、前記1重書き情報は前記揮発共有メモリ面にのみ書き込むことを特徴とするディスク制御装置を提供する。上記第2の観点によるディスク制御装置では、例えば、消失すると顧客データの破壊が発生するような重要情報は、不揮発共有メモリ面と揮発共有メモリ面の両面に2重書きし、

消失しても再構築可能で顧客データに影響を与えないような制御情報は、揮発共有メモリ面のみに1重書きするという運用が可能になる。そして、一般に、2重書きすべき情報は1重書きする情報よりも少ないため、不揮発共有メモリ面の容量を揮発共有メモリ面の合計容量よりも小さく出来る。従って、不揮発半導体メモリを採用して、バッテリーバックアップそのものを無くすることも可能となる。

【0009】第3の観点では、本発明は、上記構成のディスク制御装置において、前記不揮発共有メモリ面が独立した2以上の電源系のいずれからでも給電可能に配置され、前記揮発共有メモリ面が前記電源系の各々に分けて配置されることを特徴とするディスク制御装置を提供する。上記第3の観点によるディスク制御装置では、電源系のいずれかがダウンしても、その電源系の揮発共有メモリ面の分担を他のダウンしていない電源系の揮発共有メモリ面に割り当てることで、運用を継続することが出来る。

【0010】第4の観点では、本発明は、上記構成のディスク制御装置において、前記不揮発共有メモリ面が、ディスクドライブを含むことを特徴とするディスク制御装置を提供する。上記第4の観点によるディスク制御装置では、2重書き情報をディスクドライブに書き込んでしまう時間だけのバッテリー容量があれば、どんな長時間の電源障害にも対応できるようになる。

【0011】

【発明の実施の形態】以下、図を参照して本発明の実施の形態を説明する。なお、これにより本発明が限定されるものではない。図1は、本発明の一実施形態に係るディスクアレイ制御装置の構成図である。このディスクアレイ制御装置10は、ホストアダプタ12、ディスクアダプタ13、キャッシュメモリ14、不揮発共有メモリ15、共有メモリバススイッチ17および多数のディスクドライブ18を具備している。

【0012】前記ディスクアレイ制御装置10は、クラスAとクラスBの2つの独立した電源系に分割されている。前記クラスA、Bの各々には、1以上の前記ホストアダプタ12、1以上の前記ディスクアダプタ13、1以上の前記キャッシュメモリ14および1以上の前記共有メモリバススイッチ17が属している。また、前記不揮発共有メモリ15は、クラスA、Bの両方に属している。つまり、前記不揮発共有メモリ15は、クラスA、Bのいずれからでも電源を供給される。また、前記キャッシュメモリ14および前記不揮発共有メモリ15は、図示されていないバッテリーによってバックアップされている。

【0013】前記ホストアダプタ12は、該ホストアダプタ12を制御するチャンネルプロセッサ19、そのチャンネルプロセッサ19の指示によりホストコンピュータHとキャッシュメモリ14間のデータ転送を行うチャンネル

インタフェース回路20およびディスクアレイ制御装置10の制御情報を記憶する揮発共有メモリ16を具備して構成される。

【0014】前記ディスクアダプタ13は、該ディスクアダプタ13を制御するディスクプロセッサ21、そのディスクプロセッサ21の指示によりディスクドライブ18とキャッシュメモリ14間のデータ転送を行うディスクインタフェース回路22およびディスクアレイ制御装置10の制御情報を記憶する揮発共有メモリ16を具備して構成される。

【0015】前記チャンネルプロセッサ19および前記ディスクプロセッサ21は、共有メモリバス23により、前記共有メモリバススイッチ17に接続されている。前記共有メモリバススイッチ17は、前記チャンネルプロセッサ19または前記ディスクプロセッサ21のアクセス要求に応じて、前記不揮発共有メモリ15または前記揮発共有メモリ16とのバス接続を行う。これにより、前記チャンネルプロセッサ19および前記ディスクプロセッサ21は、任意の前記不揮発共有メモリ15または前記揮発共有メモリ16にアクセスできる。

【0016】前記ホストアダプタ12のチャンネルインタフェース回路20および前記ディスクアダプタ13のディスクインタフェース回路22は、図示されていないデータバスで、前記キャッシュメモリ14と接続されている。これにより、前記チャンネルインタフェース回路20および前記ディスクインタフェース回路22は、任意の前記キャッシュメモリ14にアクセスできる。

【0017】図2は、前記不揮発共有メモリ15および揮発共有メモリ16のエリア分割と情報の書き込み方を示している。前記不揮発共有メモリ15は、2重書きエリアのみ有している。その2重書きエリアは、構成情報エリア30と、ライトペンディング情報エリア31と、ライトコマンド通信エリア32とからなる。前記揮発共有メモリ16は、2重書きエリアおよび1重書きエリアを有している。その2重書きエリアは、構成情報エリア30と、ライトペンディング情報エリア31と、ライトコマンド通信エリア32とからなる。また、1重書きエリアは、リードコマンド通信エリア33と、キャッシュディレクトリエリア34とからなる。

【0018】前記構成情報エリア30は、2重書きエリアであり、不揮発共有メモリ15および揮発共有メモリ16の両方に同一の内容が書き込まれる。書き込まれる内容は、ディスクアレイ制御装置10のハード構成等の情報であり、例えばデータがどの物理ドライブに記憶されているか等の情報である。前記ライトペンディング情報エリア31は、2重書きエリアであり、不揮発共有メモリ15および揮発共有メモリ16の両方に同一の内容が書き込まれる。書き込まれる内容は、ホストコンピュータHからキャッシュメモリ14に書き込まれたデータであってディスクドライブ18に未反映のデータがキャ

ッシュメモリ14上のどの位置にどれだけあるかの情報である。前記ライトコマンド通信エリア32は、2重書きエリアであり、不揮発共有メモリ15および揮発共有メモリ16の両方に同一の内容が書き込まれる。書き込まれる内容は、ホストコンピュータHからデータのライト要求があった場合の処理においてチャンネルプロセッサ19とディスクプロセッサ21が行う通信の情報である。

【0019】前記リードコマンド通信エリア33は、1重書きエリアであり、揮発共有メモリ16にのみ内容が書き込まれる。書き込まれる内容は、ホストコンピュータHからデータリード要求があった場合の処理においてチャンネルプロセッサ19とディスクプロセッサ21が行う通信の情報である。前記キャッシュディレクトリエリア34は、1重書きエリアであり、揮発共有メモリ16にのみ内容が書き込まれる。書き込まれる内容は、キャッシュメモリ14上のどの位置にどのデータが記憶されているかを示す情報である。この情報は、リードデータのヒット／ミス判定を行うのに使用する。

【0020】図3は、ホストコンピュータHからデータのリード要求があった場合のディスクアレイ制御装置10の動作を示すフロー図である。ステップ40では、ホストアダプタ12のチャンネルプロセッサ19は、ホストコンピュータHからのリード要求を受け取り、要求データがキャッシュメモリ14上に存在するかどうかを確認するために、共有メモリバススイッチ17を介して、揮発共有メモリ16のキャッシュディレクトリエリア34を参照する。ステップ40aでは、要求データがキャッシュメモリ14上にない場合はステップ41へ進み、要求データがキャッシュメモリ14上にある場合はステップ46へ進む。

【0021】ステップ41では、チャンネルプロセッサ19は、揮発共有メモリ16の構成情報エリア30を参照し、要求データをディスクドライブ18より読み出しキャッシュメモリ14に書き込ませるディスクプロセッサ21を決定する。ステップ42では、チャンネルプロセッサ19は、決定したディスクプロセッサ21に対応した揮発共有メモリ16のリードコマンド通信エリア33に、キャッシュメモリ14へのデータ書き込みアドレスを指定したデータ読み出しコマンドを書き込む。ステップ43では、自分に対応する揮発共有メモリ16のリードコマンド通信エリア33をポーリングしているディスクプロセッサ21は、書き込まれた前記データ読み出しコマンドを読み取ると、ディスクドライブ18より要求データを読み出し、キャッシュメモリ14の指定されたデータ書き込みアドレスに書き込む。ステップ44では、ディスクプロセッサ21は、揮発共有メモリ16のキャッシュディレクトリエリア34のキャッシュディレクトリ情報を更新する。ステップ45では、ディスクプロセッサ21は、データ読み出しコマンドを発行したチ

ャンネルプロセッサ19に対応した揮発共有メモリ16のリードコマンド通信エリア33に、実行完了フラグを書き込む。自分に対応する揮発共有メモリ16のリードコマンド通信エリア33をポーリングしているチャンネルプロセッサ19が前記書き込まれた実行完了フラグを読み取ると、ステップ46へ進む。

【0022】ステップ46では、チャンネルプロセッサ19は、要求データをキャッシュメモリ14より読み出して、ホストコンピュータHに転送する。また、データ読み出しコマンドを揮発共有メモリ16のリードコマンド通信エリア33に書き込んでいたなら、該データ読み出しコマンドを消去する。そして、読み出し動作を終了する。

【0023】以上のとおり、ホストコンピュータHからデータのリード要求があった場合、チャンネルプロセッサ19とディスクプロセッサ21は、揮発共有メモリ16のみを使用してリード動作の制御を行う。このため、リードI/O性能は、多面化された揮発共有メモリ16のアクセス性能に支配され、不揮発共有メモリ15のアクセス性能に関係なく、十分なものが得られる。

【0024】図4は、ホストコンピュータHからデータのライト要求があった場合のディスクアレイ制御装置10の動作を示すフロー図である。ステップ50では、チャンネルインタフェース回路12のチャンネルプロセッサ19は、ホストコンピュータHからのライト要求を受け取り、共有メモリバススイッチ17を介して、揮発共有メモリ16の構成情報エリア30を参照し、ライトデータをディスクドライブ18に書き込ませるディスクプロセッサ21を決定する。ステップ51では、チャンネルプロセッサ19は、キャッシュメモリ14上にライトデータを書き込む。ステップ52では、チャンネルプロセッサ19は、決定したディスクプロセッサ21に対応した揮発共有メモリ16のライトペンディング情報エリア31に、ディスクドライブに未反映のライトデータがキャッシュメモリ14のどのアドレスに存在するかを書き込む。続いて、不揮発共有メモリ15のライトペンディング情報エリア31にも、同じ内容を書き込む。ステップ53では、チャンネルプロセッサ19は、ホストコンピュータHにデータライト完了の報告を行う。

【0025】ステップ54では、チャンネルプロセッサ19は、決定したディスクプロセッサ21に対応した揮発共有メモリ16のライトコマンド通信エリア32に、キャッシュメモリ14上のライトデータのアドレスを指定したデータ書き込みコマンドを書き込む。続いて、不揮発共有メモリ15のライトコマンド通信エリア32にも、同じ内容を書き込む。ステップ55では、自分に対応する揮発共有メモリ16のライトコマンド通信エリア32をポーリングしているディスクプロセッサ21は、書き込まれた前記データ書き込みコマンドを読み取ると、ライトペンディング情報エリア31に書き込まれた

キャッシュメモリ14のアドレスにあるライトデータを、ディスクドライブ18に書き込む。ステップ56では、ディスクプロセッサ21は、揮発共有メモリ16のキャッシュディレクトリエリア34のキャッシュディレクトリ情報を更新する。ステップ57では、ディスクプロセッサ21は、データ書き込みコマンドを発行したチャンネルプロセッサ19に対応した揮発共有メモリ16のライトコマンド通信エリア32に、実行完了フラグを書き込む。ステップ58では、自分に対応する揮発共有メモリ16のライトコマンド通信エリア32をポーリングしているチャンネルプロセッサ19は、前記書き込まれた実行完了フラグを読み取ると、揮発共有メモリ16のライトペンディング情報エリア31に書き込んでいた前記アドレスおよびライトコマンド通信エリア32に書き込んでいた前記データ書き込みコマンドを消去する。続いて、不揮発共有メモリ15のライトペンディング情報エリア31に書き込んでいた前記アドレスおよびライトコマンド通信エリア32に書き込んでいた前記データ書き込みコマンドも消去する。そして、ライト動作を終了する。

【0026】以上のとおり、ホストコンピュータHからデータのライト要求があった場合、ライトペンディング情報エリア31へのアドレス書き込み及びライトコマンド通信エリア32へのコマンド書き込みのみ不揮発共有メモリ15へのアクセスが発生し、その他の共有メモリアクセスは、すべて揮発共有メモリ16に対して行われる。このため、ライトI/O性能も、多面化された揮発共有メモリ16のアクセス性能に支配され、不揮発共有メモリ15のアクセス性能に関係なく、十分なものが得られる。

【0027】以上から明らかなように、データのリード/ライトいずれの動作でも、共有メモリのアクセスは、揮発共有メモリ16に対するアクセスが大半を占めることとなる。このため、I/O性能は、揮発共有メモリ16のアクセス性能が支配的になり、揮発共有メモリ16の多重度を上げることで、ディスクアレイ制御装置10のI/O性能の向上が可能となる。

【0028】図5は、電源障害が発生した場合のディスクアレイ制御装置10の動作を示した説明図である。電源障害には、片側のクラスタのみで障害が生じる片クラスタ電源障害70と、両方のクラスタで障害が生じる停電(両クラスタ電断)71とがある。

【0029】片クラスタ電源障害70の場合、非障害発生クラスタ72のチャンネルプロセッサ19は、自身が障害発生クラスタ73のディスクプロセッサ21に対してコマンドを発行している場合は、非障害発生クラスタ72のディスクプロセッサ21にコマンドを発行し直すことで、非障害発生クラスタ72のみを使用して動作を継続する。

【0030】片クラスタ電源障害70の場合の障害発生

クラスタ73や停電71の場合の障害発生クラスタ73(全クラスタ)は、障害発生中は、動作を停止する。障害発生がリード動作中(75)なら、ディスクドライブ18上のデータを破壊することがないため、ディスクアレイ制御装置10では特別な処理を行わない。ホストコンピュータH側でリードデータを受けたか判断し、リードデータを受けていないのであれば、障害復旧後に、ホストコンピュータHがデータのリード要求を再発行すれば良い。障害発生がライト動作中(76)なら、ライトデータがディスクドライブ18に反映されていない状態で、ホストコンピュータHにライト完了が報告されている虞れがある。このため、障害復旧後に、チャンネルプロセッサ19とディスクプロセッサ21は、不揮発共有メモリ15のライトペンディング情報エリア31とライトコマンド通信エリア32とを参照して、ライトペンディングになっているデータを全てディスクドライブ18に書き込む。これにより、ホストコンピュータH側の認識とディスクドライブ18のデータ状態とを一致させることができ、電源障害によるデータの消失を防止することが出来る。

【0031】図6は、前記不揮発共有メモリ15の第1例の構成図である。この第1例の不揮発共有メモリ15は、各クラスタに対応した複数のメモリコントロール回路81、共有メモリ情報を記憶する揮発メモリ82、通常給電系84からの供給電圧の低下を監視する電圧低下検出回路83および通常給電系84とバッテリー給電系85のいずれから不揮発共有メモリ系86(前記メモリコントロール回路81および前記揮発メモリ82を含む)に給電するかを選択的に切り替える電力供給切り替え回路87を具備して構成される。

【0032】前記メモリコントロール回路81は、共有メモリバス23経由で到着するチャンネルプロセッサ19またはディスクプロセッサ21からの指示に従い、揮発メモリ82に対するデータのリード/ライトを行う。また、前記複数のメモリコントロール回路81は、互いに通信線88で接続されており、揮発メモリ82上の同一アドレスに対するアクセスについては、衝突を避けるように互いに調停を行う。

【0033】前記電圧低下検出回路83は、通常給電系84の電圧を常に監視しており、通常給電系84の電圧低下を検出すると、制御線89を介して、前記電力供給切り替え回路87を制御し、不揮発共有メモリ系86への給電源をバッテリー給電系85に切り替え、揮発メモリ82上のデータを電断による消失から保護する。これにより、不揮発共有メモリ系86へのバッテリーバックアップが継続可能な間、電断から共有メモリ82のデータを保護することが出来る。このため、バックアップ用バッテリーの容量は、揮発メモリ82の容量と必要なバックアップ時間とから決定する必要がある。

【0034】なお、前記通常給電系84は、クラスタA

の電源系とクラスタBの電源系の両方から給電可能になっている。

【0035】図7は、前記不揮発共有メモリ15の第2例の構成図である。この第2例の不揮発共有メモリ15は、各クラスタに対応した複数のメモリコントロール回路91、共有メモリ情報を記憶する揮発メモリ92、ディスクコントローラ93、ディステージコントローラ94、ディスクドライブ95、通常給電系97の供給電圧の低下を監視する電圧低下検出回路96および通常給電系97とバッテリー給電系98のいずれから不揮発共有メモリ系99（前記メモリコントロール回路91、前記揮発メモリ92、前記ディスクコントローラ93、前記ディステージコントローラ94および前記ディスクドライブ95を含む）に給電するかを選択的に切り替える電力供給切り替え回路100を具備して構成される。

【0036】前記メモリコントロール回路91は、共有メモリバス23経由で到着するチャネルプロセッサ19またはディスクプロセッサ21からの指示に従い、揮発メモリ92に対するデータのリード／ライトを行う。また、前記複数のメモリコントロール回路91は、互いに通信線101で接続されており、揮発メモリ92上の同一アドレスに対するアクセスについては、衝突を避けるように互いに調停を行う。

【0037】前記電圧低下検出回路96は、通常給電系97の電圧を常に監視しており、通常給電系97の電圧低下を検出すると、制御線102を介して、前記電力供給切り替え回路100を制御し、不揮発共有メモリ系99への給電源をバッテリー給電系98に切り替え、揮発メモリ92上のデータを電断による消失から保護する。同時に、前記電圧低下検出回路96は、通報線104を介して、ディステージコントローラ94に対し、電断発生を通知する。この通知を受けたディステージコントローラ94は、ディスクコントローラ93を制御し、揮発メモリ92のデータを前記ディスクドライブ95に全てコピーする。これにより、バッテリー容量に関係なく、長期の電断からデータを保護することが出来る。このため、バックアップ用バッテリーの容量は、揮発共有メモリ92のデータをディスクドライブ95に書き込むのに必要な時間だけ電源供給が可能のように決定すれば良い。例えば、揮発メモリ92の容量が2GB、ディスクドライブ95へのデータ転送速度が10MB/sであれば、最低200秒間、電力が供給できれば良い。

【0038】前記電圧低下検出回路96は、通常給電系97が障害復旧したことを検出すると、前記不揮発共有メモリ系99に対する給電源を通常給電系97に切り替える。また、通報線104を介して、前記ディステージコントローラ94に障害復旧を通知する。この通知を受けたディステージコントローラ94は、ディスクコントローラ93を制御して、ディスクドライブ95上にコピーしていたデータを揮発メモリ92に戻し、チャネルブ

ロセッサ19またはディスクプロセッサ21が共有メモリ情報にアクセスできるようにする。

【0039】なお、前記通常給電系97は、クラスタAの電源系とクラスタBの電源系の両方から給電可能になっている。また、ディスクドライブ95の代わりに、他の不揮発記憶装置（光ディスクなど）や不揮発メモリ（フラッシュメモリやMagneticRAM）を使用しても良い。

【0040】上記では、電源障害を想定して説明したが、揮発共有メモリ16のいずれか自身に障害を生じた場合でも、該障害を生じた揮発共有メモリ16の分担を他の障害のない揮発共有メモリ16に割り当てることで、運用を継続することが出来る。

【0041】

【発明の効果】本発明のディスク制御装置によれば、共有メモリの多面化により高いI/O性能を得られると共に不揮発共有メモリ面を備えることで電源障害にも対応できるようになる。また、共有メモリの一部だけを不揮発共有メモリ面とすることで、回路構成やコスト上の困難が少なくなり、実施が容易になる。

【図面の簡単な説明】

【図1】本発明の一実施形態にかかるディスクアレイ制御装置を示す構成図である。

【図2】共有メモリのエリア分割および各エリアに書き込まれる情報の分類を示した説明図である。

【図3】データリード動作時のディスクアレイ制御装置の動作を示すフロー図である。

【図4】、データライト動作時のディスクアレイ制御装置の動作を示すフロー図である。

【図5】電源障害発生後のディスクアレイ制御装置のリカバリ方法を示した説明図である。

【図6】不揮発共有メモリの第1実施例を示す構成図である。

【図7】不揮発共有メモリの第2実施例を示す構成図である。

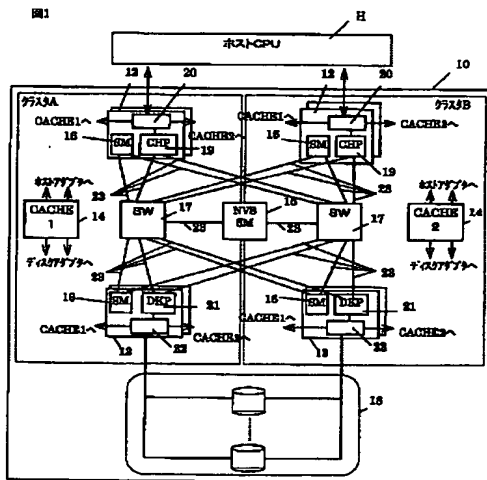
【図8】従来のディスクアレイ制御装置の一例を示す構成図である。

【符号の説明】

- |    |              |
|----|--------------|
| 1  | ディスク制御装置     |
| 2  | ホストインターフェース部 |
| 3  | ディスクドライブ     |
| 4  | ディスクインタフェース部 |
| 5  | キャッシュメモリ     |
| 6  | 共有メモリ        |
| 7  | データ用共通バス     |
| 8  | 制御情報用共通バス    |
| 10 | ディスクアレイ制御装置  |
| 12 | ホストアダプタ      |
| 13 | ディスクアダプタ     |
| 14 | キャッシュメモリ     |

- 11
- 15 不揮発共有メモリ
- 16 揮発共有メモリ
- 17 共有メモリバススイッチ
- 18 ディスクドライブ
- 19 チャンネルプロセッサ
- 20 チャンネルインタフェース回路
- 21 ディスクプロセッサ
- 22 ディスクインタフェース回路
- 23 共有メモリバス
- 30 構成情報エリア
- 31 ライトペンディング情報エリア
- 32 ライトコマンド通信エリア
- 33 リードコマンド通信エリア
- 34 キャッシュディレクトリエリア
- 81 メモリコントロール回路
- 82 揮発メモリ
- 82 共有メモリ

【図1】



【図5】

CHPのクラス		障害発生クラス70		非障害発生クラス72	
障害発生時のCHP動作		リード70	ライト70	リード	ライト
障害内容	片クラス電源障害70	障害発生後に、片クラス電源が復帰する。	障害発生後に、CHPがライトペンディング状態となり、障害発生時に、ライトペンディングになっているデータの書き込みを行う。	障害が発生したクラスのDISKメモリを、目下書き込み中のデータを保持する。	
	停電(両クラス電源)71	同上	同上		

- 12
- \* 83, 96 電圧低下検出回路
- 84, 97 通常給電系
- 85, 98 バッテリ給電系
- 86 不揮発共有メモリ系
- 87, 100 電力供給切り替え回路
- 88, 101 通信線
- 89, 102 制御線
- 91 メモリコントロール回路
- 92 揮発メモリ
- 10 92 揮発共有メモリ
- 93 ディスクコントローラ
- 94 ディステージコントローラ
- 95 ディスクドライブ
- 99 不揮発共有メモリ系
- 104 通報線
- H ホストコンピュータ
- \* MP プロセッサ

【図2】

不揮発共有メモリ系 (10)	揮発共有メモリ系 (10)	分類
構成情報エリア 30	構成情報エリア 30	
ライトペンディング情報エリア 31	ライトペンディング情報エリア 31	2重書きエリア
ライトコマンド通信エリア 32	ライトコマンド通信エリア 32	
-	リードコマンド通信エリア 33	1重書きエリア
-	キャッシュディレクトリエリア 34	

【図6】

